A VALIDATION OF STRmix[™] FOR FORENSIC CASEWORK

THESIS

Presented to the Graduate Council of the

Graduate School of Biomedical Sciences

University of North Texas

Health Science Center at Fort Worth

in Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

By

Allison Conway, B.S.

Fort Worth, Texas

May 2017

ACKNOWLEDGEMENTS

I would first like to thank my committee for support and encouragement: my major professor Dr. Bruce Budowle, Dr. Joseph Warren, Dr. Patricia Gwirtz, and Dr. Robert Mallet.

I would also like to acknowledge Harris County Institute of Forensic Sciences for hosting me and supplying all materials necessary for this project, as well as opportunities to observe the processing of DNA evidence at all stages of an investigation. My sincere thanks also goes to those individuals at HCIFS who taught, guided, assisted, and provided feedback, including but not limited to Katie Welch, Mike Donley, and Justin Foster.

Lastly, none of this work would have been possible without the financial and emotional support from my parents, George and Dorothy Conway, and my fiancé Dylan Ondarza.

TABLE OF CONTENTS

IST OF TABLES iv
IST OF ILLUSTRATIONS
hapter
I. INTRODUCTION AND BACKGROUND 1
II. MATERIALS AND METHODS
III. RESULTS
IV. DISCUSSION
V. CONCLUSIONS
EFERENCES

LIST OF TABLES

Table 1. Suggested studies listed in the SWGDAM Guidelines for Probabilistic Genotyping	
(2015). Highlighted standards indicated experiments performed in this report	7
Table 2. Samples and hypotheses used for Alternate Propositions 1	2
Table 3. LR calculations in Excel and by STRmix, with and without theta (θ) 2	2
Table 4. LR and Log(LR) results for five high copy samples	3

LIST OF ILLUSTRATIONS

Figure 1. Sample processing from extraction to analysis
Figure 2. A drop-in peak (blue) was added at a locus where the minor contributor (red) was
masked by the major contributor (black) 14
Figure 3. Peak height correlation of initial single source profile data
Figure 4. Stutter correlation plot from initial single source profile data
Figure 5. Heterozygote balance based on loci with heterozygous peaks in the initial single source
data
Figure 6. Log(Likelihood) progression across iterations
Figure 7. Log(LR)s from five single source profile dilutions
Figure 8. Log(LR) results from the single source sample 1 dilution series
Figure 9. Log(LR) results from 2, 3, and 4 person mixtures. Major contributors are on the left,
minor contributors on the right
Figure 10. Log(LR)s of two person mixtures, where $H_1 = [Major, Minor]$ and
Figure 11. Log(LR)s of three person mixtures, where $H_1 = [Major, minor, U]$ and
Figure 12. Log(LR)s of three person mixtures, where $H_1 = [Major, minor, minor]$ and
Figure 13. Log(LR)s of four person mixtures, where $H_1 = [Major, minor, U, U]$ and
Figure 14. Log(LR)s of four person mixtures, where $H_1 = [Major, minor, minor, U]$ and
Figure 15. Log(LR)s of four person mixtures, where $H_1 = [Major, minor, minor]$ and $H_2 =$
[Major, minor, minor, U]

Figure 16. Log(LR) results for true contributors (blue) and non-contributors (pink) at the correct
number of contributors. All samples amplified at 500 pg
Figure 17. Log(LR) results for true contributors (blue) and non-contributors (pink) at one greater
than the correct number of contributors
Figure 18. Log(LR) results for true contributors (blue) and non-contributors (pink) at the correct
number of contributors. Both major and minor contributors of low quantity three and four
person mixtures were tested
Figure 19. Log(LR) results for true contributors (blue) and non-contributors (pink) at one less
than the correct number of contributors
Figure 20. Log(LR) results for single source profiles with drop-in peaks added at either a
homozygous or heterozygous locus
Figure 21. Log(LR) results for two person mixtures with drop-in peaks, tested under two
different hypotheses
Figure 22. Log(LR) results for increased back stutter in single source samples up to 30% of the
parent peak
Figure 23. Log(LR) results for single source profiles with increased forward stutter to 10% and
15% of the parent peak
Figure 24. Change in Log(LR) with increasing forward stutter. Log(LR) of profile with
unenhanced forward stutter outlined in black
Figure 25. Heterozygote balance (Log(Hb)) as a function of average peak height of a
heterozygous allele pair
Figure 26. Effect on Log(LR) of artefactual peaks in single source profiles
Figure 27. Single source Total LR and HPD results in ten replicates

Figure 28. Two person mixture Total LR and HPD results in ten replicates
Figure 29. Three person mixture Total LR and HPD results in ten replicates
Figure 30. Four person mixture Total LR and HPD results in ten replicates
Figure 31. Single source HPD results at increasing numbers of iterations, replicated five times
each. Run time in minutes shown on the right
Figure 32. Two person mixture HPD results at increasing numbers of iterations, replicated five
times each. Run time in minutes shown on the right
Figure 33. Three person mixture HPD results at increasing numbers of iterations, replicated five
times each. Run time in minutes shown on the right
Figure 34. Four person mixture HPD results at increasing numbers of iterations, replicated five
times each. Run time in minutes shown on the right
Figure 35. Log(RMP) and Log(LR) from STRmix are shown against the amount of input DNA
Figure 36. Log(CPI) and Log(LR) of major and minor contributors calculated by STRmix for six
two-person mixtures are shown against the amount of total input DNA
Figure 37. Log(CPI) and Log(LR) of major and minor contributors calculated by STRmix for
three three-person mixtures

CHAPTER I

INTRODUCTION AND BACKGROUND

In forensic DNA analysis, mixture profiles, i.e., those profiles with two or more contributors, can be difficult to interpret. When there are characteristics such as stutter peaks, overlapping alleles, varying proportions of contributors, and allele dropout, the individual profiles may be difficult to resolve. Current methods of analysis employ a binary system which involves the use of thresholds to determine whether data can be used in a calculation to convey the significance of the observation. If data are below the established threshold, a statistical calculation is not performed. In order to better analyze these mixtures, software packages have been developed to mathematically determine the probability of observing an evidentiary profile. These software packages use probabilistic genotyping methods to assign statistical weight to various scenarios. Like any other method in forensic casework, it must be tested thoroughly to define its strengths and limitations. Probabilistic genotyping, utilized by STRmixTM (Institute of Environmental Science and Research, Porirua, New Zealand), is a tool which can be used to analyze complex mixtures, if properly validated.

As DNA analysis techniques have developed over the years, they have become increasingly more sophisticated. New extraction techniques, improved multiplex amplification protocols, and capillary electrophoresis instrumentation allow analysis of low quantity samples, e.g., "touch" DNA (1). These samples typically are more complex and consist of mixed profiles, the interpretation of which is challenging. Interpretation could be greatly aided by an appropriately more sophisticated biostatistical system that makes better use of the data contained within a DNA profile (2). While current methods of mixture interpretation have been validated, they often are insufficient for the level of complexity some profiles exhibit. The misusage of current methods applied to complex mixtures can result in a misleading or an incorrect statistic describing the weight of the genetic evidence (3). In order to make the best use of all information available, forensic laboratories should make efforts to validate probabilistic genotyping software. Such software can deconvolve mixed profiles, compare reference profiles, consider more than one electropherogram in analysis, and calculate statistics on the likelihood of observing a profile, all in a standardized system of analysis (4). These features reduce ambiguity that may have been previously present in mixture analysis, leading to a higher level of confidence in the results, especially in complex profiles.

A DNA profile is the genetic data obtained from biological evidence. Using Short Tandem Repeats (STRs), i.e., portions of DNA which have repeating sequences, a genotype can be observed as the number of repeated units. Each person has two alleles per locus, one from each parent. Thus, a person can have two alleles with same number of repeats (homozygous) or different numbers of repeats (heterozygous). When measured by capillary electrophoresis, each allele is separated by its size and expressed as a peak. A locus will have one peak if the person is homozygous, or two peaks if the person is heterozygous. The results from all loci tested are called a profile.

Generally, a DNA profile is characterized as a mixture if "three or more alleles are present at one or more loci and/or the peak height ratios between a pair of allelic peaks for one or more loci are below the empirically determined appropriate threshold for heterozygous peak height ratio(s)" (5). The presence of more than two true alleles indicates DNA from at least two people, except in the very rare cases of triallelic patterns. The heterozygous peak height ratio (PHR) threshold is determined through internal validation studies at each lab. During a validation study, the average heterozygote PHR is calculated and the threshold is determined from the average. If a profile contains loci with imbalanced peaks (peaks which do not fall within the threshold PHR), it could mean that there is more than one person contributing to the mixture profile. Ultimately, the entire profile must be analyzed with these characteristics in mind to determine if there is more than one individual represented.

Once a profile has been characterized as a mixture, the next step is to attempt to separate contributors, if possible. Based on the relative peak heights and the analyst's determination of the number of contributors, a major contributor may be discernable. If this is possible, the discernable profile may be treated similarly to that of a single source profile, and the Random Match Probability (RMP) can be used to calculate the rarity of that major contributor (3). If the situation is appropriate, a known contributor may be subtracted from the mixture, leaving only obligate alleles of an unknown person. In these cases, a modified RMP is calculated, using the 2p correction for loci which are suspected to be heterozygous but have only one allele expressed (3). If only one allele is present, it is possible that the sister allele is masked by the known contributor or has dropped out. Allelic dropout is the absence of data due to insufficient input DNA (6). Due to its inclusion of loci with possible dropout or allele masking, the RMP produces a stronger, more discriminating statistic than mixture analysis calculations do. However, in some cases, there may not be discernable major and minor profiles. This inability to resolve the contributors makes analysis more difficult and ambiguous.

One of the simplest statistics in mixture analysis is the Combined Probability of Inclusion (CPI). A CPI estimates the portion of a population which can be included as potential contributors

of the profile. The CPI is calculated by determining the frequency of the alleles at each locus in the profile, adding the frequencies of all possible alleles at that locus, squaring the sum, and then multiplying the values for all loci together for a frequency of the alleles in the mixture (3). Loci in which alleles may have dropped out are not included in the CPI calculation, since there is no way to include the frequency of an unknown allele. It is a highly conservative estimate, as it ignores any suspect reference profiles and includes all allele combinations, as opposed to only those which can appropriately explain the profile. For example, even if there is a distinguishable major contributor, the CPI calculation does not include that information, and instead treats all allele combinations as equally likely.

Another statistical approach which can be applied to mixture analysis is a Likelihood Ratio (LR). A basic LR equation is as follows:

$$LR = \frac{\Pr(E|H_1)}{\Pr(E|H_2)}$$

where H_1 and H_2 are two competing hypotheses (6). The numerator represents the probability of observing the evidence E under H_1 , and the denominator represents the probability of observing E under H_2 . If the LR is >1, H_1 is more likely, and if it is <1, H_2 is more likely. By using a LR, more data from the profile can be included, such as likely genotypes of contributors, and the hypotheses are specific to the case at hand.

Analysis of a DNA profile can be one of three models – binary, semicontinuous, or fully continuous. Binary methods of analysis are the simplest models. They essentially deliver a "yes" or "no" – yes the evidentiary data can be used for analysis, or no the data cannot. This decision is dependent upon validation of stochastic and analytical thresholds which analysts use to include or exclude peaks in the statistical calculation. An analytical threshold determines the height a peak must be to unequivocally declare it an allele, as opposed to noise. A stochastic threshold is the

height a peak must be such that its sister allele has not dropped out. Both of these thresholds are typically calculated in a laboratory's validation of amplification and analysis methods. CPI is an example of a binary model of statistical analysis. It does not accommodate the possibility of allele dropout and ignores peak heights of alleles that are deemed to be true. The CPI is not inappropriate, but does not use all the information available in a profile, and is therefore much less specific (3).

Semicontinuous methods of analysis take into account the possibility of dropout, but not relative peak heights. To account for dropout, a fixed allele dropout rate is applied to all loci regardless of peak height. If a locus has more intense peaks (above the stochastic threshold) there is a very low chance that any dropout has occurred. However, the dropout rate applied to such a locus is the same as the rate applied to a locus with very low peak heights. It is necessary for an analyst to determine whether an individual is included or excluded based on the peak heights (7). Therefore, a semicontinuous method essentially is a binary system that includes the probability of allele dropout.

A fully continuous model uses peak height information and can account for the possibility of allele drop-in (e.g., contamination) and drop out and gives each scenario a calculated weight of probability in the form of a LR. Drop-in describes the unexplained and unreproducible presence of an allelic peak in a profile (4). Bille et al. (7) compared match statistics produced from CPI, RMP, a semicontinuous model called LabRetriever, and STRmix, a fully continuous model. The results indicated that STRmix and RMP consistently outperformed CPI and LabRetriever when it came to the more discriminating value or statistic (7). In mixtures which had a clear major/minor and the minor was present in small amounts, the CPI could not accommodate several loci which had alleles below the stochastic threshold, and LabRetriever could not make exclusions based on peak height. However, since STRmix does not use thresholds to evaluate alleles, but instead the balance of peak heights, it was able to interpret mixtures whether the two contributors were present in similar or unbalanced concentrations. Therefore, it was concluded that STRmix made the most effective use of the data present (7).

STRmix is able to simulate many different scenarios using the mathematical method Markov Chain Monte Carlo (MCMC). This method evaluates the peak height and attempts to find a scenario or scenarios (genotype combination and proportion of each contributor) with the best fit to the profile (8). The Markov chain allows a memoryless progression of sampling, like walking, where each step that is closer to the data is accepted and each step that is further from the data is rejected (9). Due to the randomness of the method, different LRs will be computed every time it is run. This variation due to sampling is expected, and the degree of deviation of computed LRs should be a consideration during validation (7).

STRmix is just one example of a fully continuous model for DNA analysis. Another model, TrueAllele®, has a similar premise and uses the same general mathematical principles. In a validation of TrueAllele, Greenspoon et al. (10) reported highly sensitive and specific results when testing true contributors and non-contributors in mixed profiles, and the Virginia Department of Forensic Science has since implemented TrueAllele in casework analysis.

Kerr (1) described the use of a validated STRmix system at the New South Wales Forensic and Analytical Science Service in Australia and claimed that the software is effective, especially in analysis of complex mixtures. Based on these results, and the need for new mixture analysis techniques, it is time to explore the opportunities that STRmix can provide.

CHAPTER II

MATERIALS AND METHODS

A validation study consists of a series of tests designed to assess a wide variety of scenarios which should determine the strengths and limitations of a system prior to its use. The types of studies conducted were chosen to comply with the Scientific Working Group on DNA Analysis Methods (SWGDAM) Guidelines for the Validation of Probabilistic Genotyping Systems (2015). The studies suggested in the SWGDAM guidelines are shown in Table 1 (11). The standards highlighted in yellow will be discussed in this report. The Implementation and Validation Guide provided with the software lists studies specific to STRmix which correspond to the SWGDAM guidelines.

Table 1. Suggested	studies listed	l in the SWGDA	M Guidelines for	Probabilistic	Genotyping
				0 /0 /0 /0 /0 /0 /0 /0 /0	

Standard	Internal Validation study description
4.1.1	Specimens with known contributors and case-type specimens that may include unknowns
4.1.2	Hypothesis testing with contributors and non-contributors
4.1.3	Variable DNA typing conditions (e.g., any variations in regular protocol to modify detection of alleles or artifacts)
4.1.4	Allelic peak height including off-scale peaks
4.1.5	Single-source specimens
4.1.6	Mixed specimens with different contributor ratios, total DNA quantities, and numbers of contributors
4.1.7	Partial profiles due to dropout, degradation, and inhibition
4.1.8	Allele drop-in
4.1.9	Forward and reverse stutter
4.1.10	Intra-locus peak height variation

(2015). Highlighted standards indicate experiments performed in this report.

4.1.11	Inter-locus peak height variation
4.1.12	Use of different data to establish in-house parameters for software
4.1.13	Sensitivity, specificity, and precision
4.1.14	Additional challenge testing, such as including non-allelic peaks
4.2	Comparison to manual interpretation

Samples were processed from extraction to analysis according to Figure 1. Samples were extracted from buccal swabs using a QIAsymphony SP (Qiagen, Hilden, Germany) automated system. The quantity of recovered DNA was determined with Applied Biosystems[™] Quantifiler® Trio (Thermo Fisher Scientific, South San Francisco, CA) on an Applied Biosystems 7500 Real Time PCR System (Thermo Fisher) and diluted to 500 pg/µL for stock solutions. To create mixture profiles, stock solutions were combined in appropriate ratios. At this point, samples were serially diluted if necessary. Samples were amplified with Applied Biosystems GlobalFiler® PCR Amplification kit (Thermo Fisher) on an Applied Biosystems GeneAmp® PCR System 9700 (Thermo Fisher). The PCR products were separated on an Applied Biosystems 3130xL Genetic Analyzer (Thermo Fisher), then analyzed and edited using GeneMapper® ID-X (Thermo Fisher). Any artefacts, or non-PCR products, were edited out of the profile, but stutter was not removed. The genotype table for each sample was exported to Microsoft Excel. Evidence profiles contained only true alleles and stutter peaks, along with corresponding sizes and heights. Reference profiles contained only true alleles, along with only sizes, and no peak height data. Once the files were formatted as the STRmix manual specified, they were imported to the software and analyzed.

A STRmix report contains several different types of results. The report presents the weights for each possible genotype at each locus, the Total LR for each population, and the 99% lower bound highest posterior density LR (HPD). A weight for a genotype will be between 0 and 1, and the sum of all the weights at a locus equals 1. The Total LR is a point estimate calculated for each locus based on the genotype weight. STRmix then calculates the 99% confidence interval of the

Total LR. The lower bound of the confidence interval is reported as the HPD, meaning that, with 99% confidence, the true LR value lies above the HPD. The developers of STRmix intended for the HPD to be conservative in order to account for the variability due to MCMC. The HPD value is used throughout this report as the LR value, unless otherwise specified.



Figure 1. Sample processing from extraction to analysis

Initial Settings

The Implementation Guide for STRmix also gives instructions on how to set up the initial parameters: stutter, analytical threshold, saturation limit, drop-in parameters, variance constant distributions for stutter and allele models, locus specific amplification efficiency, and population settings (12). These parameters are based on lab-specific empirical data. The stutter ratios are calculated *per allele*, instead of generally per locus as has been traditionally done. By calculating

stutter ratio per allele within each locus, the overall stutter ratio for the locus can be described as a linear equation through a regression analysis. The linear equation results in a more accurate calculation of the estimated stutter peak height than a static value which applies to all alleles regardless of size.

The analytical threshold and capillary electrophoresis instrument's saturation limit had been determined through the lab's validation of the amplification kit, Applied Biosystems GlobalFiler PCR Amplification kit, with a different analytical threshold per dye channel. Drop-in parameters were determined from actual instances of drop-in from previously run samples. Variance constant distributions and locus specific amplification efficiencies were calculated by Model Maker, a program within STRmix, from single source profile data. Population data were based on FBI allele frequencies for African American, Caucasian, and Southwest Hispanic populations, with a theta value of 0.01, which is typically used in the American forensic community (13).

Single Source

The validation began with a check of the single source LR calculations. For single source profiles (without dropout), the software should result in a consistent Total LR, since it does not have to consider multiple genotype possibilities. This point estimate should be able to be reproduced by a manual calculation based on the Balding and Nichols formulae, below (13).

Homozygote, [i, i]:
$$\frac{|3\theta+(1-\theta)p_i||2\theta+(1-\theta)p_i|}{(1+\theta)(1+2\theta)}$$
Heterozygote, [i, j]:
$$\frac{2[\theta+(1-\theta)p_i][\theta+(1-\theta)p_j]}{(1+\theta)(1+2\theta)}$$

-> 1- - /

Five single source profiles were run on STRmix. The weights of each genotype were noted, and the Total LRs were compared against manual calculations.

High Copy

High concentration samples were tested. These samples were amplified at 4 ng, which is much higher than the ideal input amount, and the goal was to obtain results which would saturate the camera of the CE instrument. When a peak is over the saturation limit, its height is not accurately measured, which could affect the accuracy of the calculation of the expected stutter peak height. Even in such cases, STRmix should still place a weight of 1 on the correct genotype, rather than considering a stutter peak as a true allele.

Mixture Weights

Two sets of two, three, and four person mixture series were created, all with the same DNA input amount of 500 pg. One set was comprised of two similar ("S") profiles, which shared alleles at many loci, while the other set was comprised of two different ("D") profiles, which did not share many alleles. These two series were designed to test how allele masking affects the LR calculation. LRs were calculated for both the major and minor contributors, separately, and at each mixture ratio (10:1, 5:1, 3:1, 2:1, and 1:1). The LR hypotheses were

H1: The mixture is made up of the person of interest (POI) and N-1 unknown persons, and

H₂: The mixture is made up of N unknown persons,

where N is the number of contributors to the mixture, and the POI is either the major or a minor contributor.

Alternate Propositions

While the Weights section tested only the major or one minor contributor, with no assumed known contributors, the Alternate Propositions study aims to test the addition of assumed known contributors to H₂. The two, three, and four person mixtures from Weights were re-analyzed under various hypotheses, listed in Table 2.

	S	D
Two	$H_1 = [VD173, VD300]$	$H_1 = [VD160, VD259]$
person	$H_2 = [VD173, U]$	$H_2 = [VD160, U]$
	$H_1 = [VD173, VD300, U]$	$H_1 = [VD160, VD259, U]$
Three	$H_2 = [VD173, U, U]$	$H_2 = [VD160, U, U]$
person	$H_1 = [VD173, VD300, VD320]$	$H_1 = [VD160, VD259, VD319]$
	$H_2 = [VD173, VD300, U]$	$H_2 = [VD160, VD259, U]$
	$H_1 = [VD173, VD300, U, U]$	$H_1 = [VD160, VD259, U, U]$
	$H_2 = [VD173, U, U, U]$	$H_2 = [VD160, U, U, U]$
Four	$H_1 = [VD173, VD300, VD320, U]$	$H_1 = [VD160, VD259, VD319, U]$
Person	$H_2 = [VD173, VD300, U, U]$	$H_2 = [VD160, VD259, U, U]$
	$H_1 = [VD173, VD300, VD320, VD85]$	$H_1 = [VD160, VD259, VD319, VD105]$
	$H_2 = [VD173, VD300, VD320, U]$	$H_2 = [VD160, VD259, VD319, U]$

Table 2. Samples and hypotheses used for Alternate Propositions¹

Number of Contributors

As discussed previously, assigning the true number of contributors can be difficult for complex mixtures, and profiles should be tested with hypotheses that entertain the wrong number of contributors to assess the effect on the LR if an estimation error is made by an analyst. Ten

¹ Sample names are VD173, VD300, VD320, VD85, VD160, VD259, VD319, and VD105 (denoted VD for validation samples). "S" and "D" refer to the similar and different sets of profiles as discussed in the Mixture Weights study. "U" designates an unknown, untested person.

single source profiles, ten two person mixtures, and ten three person mixtures were analyzed as two, three, and four person mixtures, respectively, to simulate overestimating of the number of contributors. The profiles were also run against a database of 200 randomly generated Caucasian profiles to test the effect of non-contributors on the LR. The resulting LRs were plotted against the lowest average peak height (APH) of all contributors of a particular profile. To simulate underestimation of contributors, 12 three person and 9 four person mixtures from the Weights and Sensitivity studies were analyzed as two and three person mixtures, respectively. These profiles were also run against the random database, and the LRs plotted against the minimum APH.

Drop-in

Testing for drop-in includes addition of an artificial drop-in peak to single source and two person mixture profiles. Drop-in peaks were added at either 70 or 100 rfu to span the drop-in threshold of 75 rfu. This threshold was determined from previous drop-in data from the lab. The peaks were added to either a homozygous or a heterozygous locus in the single source profiles and to a locus in the two person mixtures where the homozygous minor contributor was masked by the heterozygous major contributor. The addition of a drop-in peak to the two person mixture is diagrammed displayed in Figure 1. The major contributor was a [17, 18], while the minor contributor (red) was a [17, 17]. A drop-in peak (blue) was added at allele 15 at either 70 or 100 rfu.



Figure 2. A drop-in peak (blue) was added at a locus where the minor contributor (red) was masked by the major contributor (black).

Samples tested included single source dilution series profiles from 500 to 16 pg, and two person mixture ratios from 10:1 to 1:1 at 500 pg (Weights study). LRs from profiles with drop-in were compared to the LR of the samples without drop-in.

Stutter

Heights of stutter peaks were increased incrementally in single source profiles, both at heterozygous and homozygous loci, and for back and forward stutter. The LRs of the artificially increased stutter peaks were plotted against the LRs of the profiles with natural stutter. Weights were checked for each run to ensure that no weight was given to any stutter peaks.

Intra-locus peak heights

Intra locus peak heights were studied as part of the initial implementation process. Single source samples were analyzed using STRmix, and Model Maker compiled the heterozygote balance data.

Challenge samples

"Challenge" testing includes addition of artefacts to profiles. Artefacts should be removed by the analyst prior to input. An artefact is any peak which does not represent a PCR product, such as pull-up. Pull-up is spectral overlap from the dyes used during capillary electrophoresis. It is more common in high quantity samples, and can be recognized and removed by an analyst. However, if there is an artefactual peak mistakenly left in the file, this study should show the effect on the LR. Since the drop-in cap is 75 rfu, artefacts of either 50, 100, 150 or 200 rfu were added to single source profiles. The 50 rfu peaks are under the threshold and may be considered drop-in, while the higher peaks are above the drop-in threshold and may be considered true peaks.

Precision

Precision testing is important for validation of STRmix, considering the inherent randomness of the MCMC process. Precision is studied here in two ways. First, STRmix was run ten times with exactly the same conditions for each of the following: a single source profile, a two person mixture, a three person mixture, and a four person mixture. These runs were set at 500,000 iterations (100,000 burn-in and 400,000 post burn-in), which is the default value. Then, the Total LR and highest posterior density (HPD) calculations were compared across runs.

The number of iterations can affect precision, especially depending on how complex the mixture is. With a lower number of iterations, the calculation may not be complete, and more variation between runs is likely. With a higher number of iterations, the calculation will probably be more precise, but the run will take longer. To test whether the default value is a good balance between precision and run time, profiles were run five times each at 50,000, 500,000, and 5,000,000 iterations (10,000, 100,000, and 1,000,000 burn-in, respectively).

CHAPTER III

RESULTS

Initial Settings

Model Maker generated data that are examined in order to assess whether the software is running as expected. The peak correlation data (Figure 3) focuses on the two alleles of a heterozygote pair. The x-axis is the log of the observed peak height (O) over the expected or calculated peak height (E) of the lower molecular weight peak. The y-axis is the log of the observed peak height over the expected peak height of the higher molecular weight peak. Ideally, O/E would be 1, indicating that both peak heights were accurately predicted by the software, and the log(1) = 0. The data points were centered around (0,0). Although there were a few points that strayed from the center mass, they were within the expected range and were similar to the example provided by the developers.



Figure 3. Peak height correlation of initial single source profile data

The stutter correlation data (Figure 4) compare stutter peak to parent allele peak. The majority of the data points were centered around (0,0), but there were some points which were skewed to the right. These results indicated a correlation between stutter and parent peak heights, which is expected due to the nature of stutter peaks and their dependency on the true allele during the PCR process.



Figure 4. Stutter correlation plot from initial single source profile data

The heterozygosity balance graph (Figure 5) presents the balance of the heterozygous peak heights in relation to their average peak height. The dotted lines represent the bounds of heterozygosity balance, and 95% of data points should reside between the lines. In our study, 96% of the data fits between the bounds. These data will be discussed further in the "Intra-locus peak heights" section.



Figure 5. Heterozygote balance based on loci with heterozygous peaks in the initial single source data

The log(likelihood) progression (Figure 6) shows the LR calculation at each cycle of calculation. According to the developers, the log(LR) should begin to plateau as the software approaches the true value. In this study, the log(LR) appears to plateau after about 60 iterations.



Figure 6. Log(Likelihood) progression across iterations

Weights

Single source results are shown in Figures 7 and 8. The LR point estimate is constant for a full profile, but begins to drop with decreasing amount of input DNA due to decreasing weight placed on the true genotype. Manual LR calculations were performed with and without theta in Excel and compared to STRmix LRs, shown in Table 3. STRmix LRs were reproducible using the appropriate equations.



Figure 7. Log(LR)s from five single source profile dilutions



Figure 8. Log(LR) results from the single source sample 1 dilution series

	$\theta = \theta$	0.01	$\theta = 0$		
	Excel STRmix		Excel	STRmix	
D3S1358	13.7549	13.7549	14.5629	14.5629	
vWA	12.3123	12.3123	14.0300	14.0300	
D16S539	5.3071	5.3071	5.3917	5.3917	
CSF1PO	8.5128	8.5128	9.3995	9.3995	
TPOX	3.2193	3.2193	3.3523	3.3523	
Yindel					
D8S1179	12.8515	12.8515	13.7653	13.7653	
D21S11	15.4766	15.4766	16.4824	16.4824	
D18S51	81.5728	81.5728	107.7359	107.7359	
DYS391					
D2S441	9.4165	9.4165	10.4808	10.4808	
D19S433	11.3429	11.3429	12.8281	12.8281	
TH01	9.6950	9.6950	10.8166	10.8166	
FGA	43.1837	43.1837	50.2808	50.2808	
D22S1045	4.3117	4.3117	4.3507	4.3507	
D5S818	70.4174	70.4174	158.0140	158.0140	
D13S317	195.0895	195.0895	287.8868	287.8868	
D7S820	19.4290	19.4290	20.9863	20.9863	
SE33	74.3090	74.3090	91.3618	91.3618	
D10S1248	10.7489	10.7489	11.3449	11.3449	
D1S1656	54.7514	54.7515	66.9058	66.9058	
D12S391	24.8562	24.8562	27.4413	27.4413	
D2S1338	124.1729	124.1729	223.4925	223.4925	
Total LR	2.78E+27	2.78E+27	1.10E+29	1.10E+29	

Table 3. Total LR calculations in Excel and by STRmix, with and without theta (θ)

High copy LR results are shown in Table 4. At some loci, the peaks were so oversaturated that they were not labelled by the genotype editing software. The peaks were observable to an analyst, but there were no allele, size or height data given by GeneMapper. These data were not included in the input file, and the loci were left blank. Due to the missing data, these loci received a weight of 0. However, all loci with sufficient data received a weight of 1.

	1	2	3	4	5	
CAU						
LR	5.10E+25	1.51E+28	5.06E+26	2.45E+25	6.76E+25	
Log(LR)	25.71	28.18	26.70	25.39	25.83	

Table 4. LR and Log(LR) results for five high copy samples

Mixture results for major and minor contributors in the "S" and "D" series are shown in Figure 9. The hypotheses for the major contributor were set up so that only the major contributor was considered in H_1 , while H_2 was composed of all unknown profiles. These propositions can be summarized for a two person mixture as below, where U is an unknown, untested profile:

$$\frac{H_1 = [Major, U]}{H_2 = [U, U]}$$

The minor contributor was tested in the same manner, so that only that contributor was tested in H_1 . These hypotheses result in a LR of the probability of observing the evidence given that the contributor in question is a donor to the mixture, rather than an unknown person.

The major contributors in all samples exhibited a similar trend, decreasing steadily as the amount of template DNA placed in the PCR for that contributor decreased. The LRs of the minor contributors of the two person mixtures and of the three person mixture "D" shared a common progression. The LRs were lower at ratios of 10:1 or 10:1:1 where the amount of minor contributor template DNA was low, then increased at ratios of 5:1, 5:1:1, 3:1, and 3:1:1 where the peak heights of the major and minor contributors were more easily distinguished. The LRs then decreased at 2:1, 2:1:1, 1:1, and 1:1:1 where the major and minor peaks were no longer distinguishable or the contributors were equivalent.

Three groups of minor contributors did not fit this trend. The three person mixture "S" LR increased steadily with increasing template, as opposed to exhibiting the curve of the other minor

contributors. This trend can be attributed to the sharing of alleles with the major contributor. At low quantity, the minor contributor is barely observable. The alleles which are shared between the contributors often do not reflect the presence of the masked minor contributor. As the amount of minor contributor DNA increases, the shared peak heights become more imbalanced and the profile is easier to deconvolve with the software.

The minor contributors of both four person mixture series did not exhibit similar trends to other minor contributors, or any discernable trend at all. This is likely due to the complexity of four person mixtures.







Figure 9. Log(LR) results from 2, 3, and 4 person mixtures. Major contributors are on the left, minor contributors on the right.

Alternate Propositions

The results of the various hypotheses in the Caucasian population as well as database searches are summarized in Figures 10-15. Log(LR) is plotted against APH of the contributor in question. True contributors are shown in blue, while non-contributors from the database are shown in pink. As the number of assumed known contributors in H₂ increases, the likelihood for false inclusions decreases.



Figure 10. Log(LR)s of two person mixtures, where H₁ = [Major, Minor] and

 $H_2 = [Major, U]$



Figure 11. Log(LR)s of three person mixtures, where H₁ = [Major, minor, U] and



Figure 12. Log(LR)s of three person mixtures, where $H_1 = [Major, minor, minor]$ and



H₂ = [Major, minor, U]

Figure 13. Log(LR)s of four person mixtures, where $H_1 = [Major, minor, U, U]$ and



Figure 14. Log(LR)s of four person mixtures, where $H_1 = [Major, minor, minor, U]$ and

 $H_2 = [Major, minor, U, U]$



Figure 15. Log(LR)s of four person mixtures, where H₁ = [Major, minor, minor]

and H₂ = [Major, minor, minor, U]

Data for single source, two person mixtures, and three person mixtures are shown in Figures 16 and 17 for the correct number of contributors and one greater than the correct number, respectively.

Data for three and four person mixtures are shown in Figures 18 and 19 for the correct number of contributors and one fewer than the correct number, respectively.

For all samples, H_1 was analyzed with only one reference, against H_2 with all unknown contributors. Each contributor to the mixture was tested in turn.

When the number of contributors is overestimated by 1, the LRs of known contributors are generally unchanged, but known non-contributors are more likely to be included. The highest LR of a non-contributor was 3417 (Log(LR) = 3.53), in a three person mixture (5-1-1 D). Upon inspection of the random profile generated, all alleles except 4 sister alleles were represented by a combination of the three true contributors. With a minimum APH of 151 rfu, STRmix gave this profile a higher LR due to a greater possibility of dropout at some loci. However, even with almost all alleles present, the LR is only 3.4×10^3 .

When the number of contributors is underestimated by 1, the LRs of both known contributors and non-contributors are typically lower. Since this portion of the study used several low quantity samples from the sensitivity study, LRs for minor contributors were already very low even with the correct number of contributors. With one less contributor in the calculation, these samples began to result in LRs less than one. Major contributors and contributors with average peak heights over ~150 rfu were not significantly affected.



Figure 16. Log(LR) results for true contributors (blue) and non-contributors (pink) at the







greater than the correct number of contributors.



Figure 18. Log(LR) results for true contributors (blue) and non-contributors (pink) at the correct number of contributors. Both major and minor contributors of low quantity three and four person mixtures were tested.





less than the correct number of contributors.

Drop-in

The results of adding a drop-in peak to a single source profile in either a heterozygous or homozygous locus are shown in Figure 20. When the drop-in peak was higher than 75 rfu, no results were obtained – either due to a software error or due to a LR of 0. When the peak greater than 75 rfu was added to homozygous locus vWA, a LR of 0 was obtained for all dilutions, as the software considered that peak to be a sister allele. When the peak was added to heterozygous locus D8S1179, an error was obtained if more than two alleles were present, and a LR of 0 was obtained if one of the true sister alleles had dropped out, as in samples 1E and 1F.

Results from addition of a drop-in peak to two person mixtures are summarized in Figure 21. When the major contributor was assumed known in H₂, there was virtually no difference in LR between the original profile and the profile with drop-in. When the major contributor was not assumed and only the minor contributor was included in H₁, there was a slight difference, especially in samples where the minor contributor was present in low quantity, e.g. a 10-1 mixture. However, as the minor contributor's peak heights increased, the LRs of the original profile and profile with drop-in converged. The true contributor was never excluded.



Figure 20. Log(LR) results for single source profiles with drop-in peaks added at either a





Figure 21. Log(LR) results for two person mixtures with drop-in peaks, tested under two

different hypotheses.

Forward and Reverse Stutter

Log(LR) results for increased stutter at both heterozygous and homozygous loci are shown in Figure 22. Log(LR) results for 10% and 15% forward stutter are shown in Figure 23, and results for all ratios tested are shown in Figure 24.

Increased stutter at one locus does not seem to significantly affect the LR of the entire profile, provided that the stutter is below the threshold of 30% of the parent peak for back stutter and 15% for forward stutter. The one exception to this was in sample 2A, where an increased back stutter peak of 20% resulted in a LR of 0. It is unclear why STRmix placed a weight of 1 on [9, 10] instead of on the true genotype [10, 10] when the stutter peak height ratio was below 30%. All other profiles with increased back stutter up to 30% resulted in LRs within 1 order of magnitude of the original profile.





the parent peak



Figure 23. Log(LR) results for single source profiles with increased forward stutter to 10%

and 15% of the parent peak



Figure 24. Change in Log(LR) with increasing forward stutter. Log(LR) of profile with

unenhanced forward stutter outlined in black.

The heterozygote balance data present as expected (Figure 25). The upper and lower bounds, which are a measure of the allele variance range expected by STRmix, encompass 96% of the data points, indicating that the allele variance value is a good estimation. The STRmix manual indicates that at least 95% of the data should be within the boundaries.



Figure 25. Heterozygote balance (Log(Hb)) as a function of average peak height of a heterozygous allele pair

Challenge Testing

The effects of adding artefacts into profiles are shown in Figure 26. In sample 1A, the artefact was in a forward stutter position of one of the two true alleles. At 50, 100, and 150 rfu, the LR was not significantly affected, and only slightly lower than the LR for the same profile with the artefact removed. The threshold for the artefactual peak was much higher for this location due

to the possibility of forward stutter. At 200 rfu, when the forward stutter ratio exceeded 16%, the calculation was not completed and an error message indicated that there were too many alleles to be attributed to one person.

In sample 2A, the artefact was added to a heterozygous locus in a non-stutter position. The LR was slightly lower when an artefact of 50 rfu was added, but still sufficiently high for a single source sample. When the artefact peak height was increased to 100 rfu, an error message indicated that there was more than one contributor present. It can be concluded that the artefact at 50 rfu was determined by the software to be drop-in, but once the height exceeded the drop-in threshold of 75 rfu, it was considered a true allele.

In sample 4A, the artefact was added to a homozygous locus. The LR of the profile with a 50 rfu artefact was slightly lower than the LR of the profile with no artefacts, which is consistent with a consideration of drop-in. When the peak height was increased to 100 rfu, the LR was 0, and the software no longer considered [16, 16] to be an allowable genotype.



Figure 26. Effect on Log(LR) of artefactual peaks in single source profiles

Precision

HPD vs. Total LR results are shown in Figures 27-30, first on a wide scale, then zoomed in. Number of iteration results are shown in Figures 31-34, with the average run times overlaid on the graphs. In single source and two and three person mixtures, the HPD and Total LR lay within one order of magnitude. All HPD values were less than all Total LR values – that is, the HPD calculations for ten replicates were never higher than the lowest Total LR calculation. This indicates that the HPD calculation is accurate in being the lower bound of the confidence interval.

In four person mixtures, the lowest HPD and the highest Total LR span two orders of magnitude, rather than one. The highest HPD calculation is also higher than the lowest Total LR value. These results indicate that results for four person mixtures are not as precise as results from profiles with one, two, and three contributors.

As the number of iterations increases, precision generally increases, as does run time. In all samples, 50,000 iterations seem to be insufficient to complete the calculations due to high variation. In single source and two person mixtures, the Log(LR) values of 500,000 and 5,000,000 iteration runs all lie within 0.2 units. Run times were also acceptable, with the longest run time at 20 minutes, for a two person mixture with 5,000,000 iterations. In three person mixtures, Log(LR) values of 500,000 iteration runs lie within 0.6 units with an average of 4 minutes of analysis, and the Log(LR) values of 5,000,000 iterations runs lie within 0.2 units with an average of 30 minutes of analysis.

Analysis times of four person mixtures sharply increased. While at 50,000 iterations, analysis took an average of just 13 minutes, results spanned across two orders of magnitude, with Log(LR)s anywhere from 4 to 6. At 500,000 iterations, the average analysis time was just over one

hour, and Log(LR) values were between about 5 and 5.6. At 5,000,000 iterations, average analysis time jumped to 404 minutes, or almost 7 hours, with Log(LR) values from 5.5 to 5.8.



Figure 27. Single source Total LR and HPD results in ten replicates



Figure 28. Two person mixture Total LR and HPD results in ten replicates



Figure 29. Three person mixture Total LR and HPD results in ten replicates



Figure 30. Four person mixture Total LR and HPD results in ten replicates



Figure 31. Single source HPD results at increasing numbers of iterations, replicated five



times each. Run time in minutes shown on the right.

Figure 32. Two person mixture HPD results at increasing numbers of iterations, replicated

five times each. Run time in minutes shown on the right.



Figure 33. Three person mixture HPD results at increasing numbers of iterations,



replicated five times each. Run time in minutes shown on the right.

Figure 34. Four person mixture HPD results at increasing numbers of iterations, replicated

five times each. Run time in minutes shown on the right.

Comparison to manual interpretation

Single source profile statistics as calculated by STRmix are compared to RMP calculations in Figure 35. The match statistics for both methods are comparable even at lower quantities. This is likely because RMP does not use apparent homozygous peaks below the stochastic threshold, but can use heterozygous peaks regardless of the peak imbalance. STRmix can use the apparent homozygous peaks even when they are below the threshold, but may place a lower weight on heterozygous peaks if they are imbalanced. Depending on how many homozygous or heterozygous loci the profile contains, STRmix could give a higher or lower statistic. However, both methods tend to give comparable results and either could be used for single source profiles.



Figure 35. Log(RMP) and Log(LR) from STRmix are shown against the amount of input DNA

Comparisons of CPI and STRmix LRs for six two-person mixtures and three three-person mixtures are shown in Figures 36 and 37, respectively. In all samples, the LRs of both the major and minor contributors were greater than the CPI, even at lower quantity samples.



Figure 36. Log(CPI) and Log(LR) of major and minor contributors calculated by STRmix

for six two-person mixtures are shown against the amount of total input DNA





for three three-person mixtures

CHAPTER IV

DISCUSSION

For each study in this validation, there were certain expectations indicating whether the results from the software would be useful for a sample type. Sample types can first be sorted by amount of input DNA, which is correlated to average peak height across a profile. When there is less input DNA, peak heights are lower and dropout is more likely to occur, resulting in a lower LR. Eventually, there comes a point at which there is not enough useful information for STRmix to use. This point typically occurs when the average peak height of a contributor is less than 150 rfu, as seen in the Alternate Propositions and Number of Contributors studies.

Samples can then be sorted by number of contributors. Samples with one or two contributors are the easiest to analyze, just as with manual interpretation. Samples with three contributors could still result in reliable statistics, depending on the amount of information present. From the Alternate Proposition study, it was clear that good results could be obtained for minor contributors in three person mixtures, especially when the average peak height was higher than 150 rfu or there was a contributor assumed known in H₂, reducing uncertainty. Precision results also indicated that three person mixtures can be reproduced within the expected degree of variance. However, samples with four contributors often proved to be too complex to provide useful results. In the Weights study, the major contributors of four person mixtures were given high LRs with trends similar to other mixtures, but the minor contributors were also highly affected by a decrease in

the number of contributors, and were unable to reliably result in reproducible HPD calculations (Precision). Samples with four contributors also take the longest to deconvolve, typically running for at least an hour per sample.

The final sample type includes samples with atypical peaks present – those studied in Dropin, Stutter, and Challenge Testing. In all of these samples, STRmix dealt with the atypical peak well, as long as it fell within the expected boundaries. For drop-in peaks, the peak must be below 75 rfu. This threshold was determined from empirical data. Below the threshold, STRmix could accommodate for drop-in and resulted in a slightly lower LR. Above the threshold, the peak is not considered to be drop-in, which affected LRs of true contributors.

In a similar manner, back stutter and forward stutter had thresholds of 30% and 15%, respectively, based on empirical data as well as developer recommendations. Below these maximum stutter ratios, there is little effect on LRs, but above the thresholds, the stutter peaks were likely to be called true peaks.

For Challenge Testing, or artefacts in the profile, STRmix does not have any function to recognize pull-up. Instead, if the pull-up peak is less than 75 rfu, it is considered drop-in and does not significantly affect the LR. If it is higher than 75 rfu, it is considered a true peak, just like drop-in.

These three studies demonstrate the importance of a thorough review of the evidentiary profile by the analyst. While drop-in can be difficult to identify, pull-up peaks should be easily recognized and removed by the analyst. Any peaks in a stutter position over 30% or 15% for back or forward stutter, respectively, should be examined to determine whether the peaks are true alleles from another contributor, or if something else is contributing to the stutter peak heights, such as pull-up. The POI should always be included or excluded by the analyst prior to beginning a

47

STRmix run. If the POI is included by the analyst, then a LR of 0 indicates a user error rather than an exclusion – the number of contributors or hypothesis setup may have been entered incorrectly, or an artefact may have been left in the profile. For this reason, it is important that all analysts using STRmix are trained to best use the system prior to conducting analyses.

When compared to RMP calculations for single source profiles, STRmix performs comparably. If the profile has more potential homozygous peaks, i.e., single peaks below the stochastic threshold, STRmix may be favored due to its ability to factor dropout into the calculation. However, both RMP and STRmix result in appropriate match statistics.

When calculating CPI of two person mixtures, some loci had to be excluded due to possible dropout. The differences in match statistics between CPI and STRmix LR are notable, particularly with the major contributor. Since STRmix can compare the possible genotypes of the profile to a reference, the major and minor contributors can be evaluated separately, leading to a dramatically increased LR of the major contributor when the minor contributor is present in trace amounts. The LR of the minor contributor was also greater than the CPI in all samples tested.

Calculation of CPI for three person mixtures is significantly more difficult, and is rarely done in casework. Three profiles were able to be analyzed with CPI, all of which were at the optimum range of DNA input at 500 pg. Again, all LRs for both major and minor contributors were greater than the CPI for those samples. Perhaps more significant is the amount of data in profiles that could not be analyzed by CPI, all of which returned results from STRmix.

CHAPTER V

CONCLUSIONS

An increase in more complex DNA profiles demands an appropriately more sophisticated method of statistical analysis. Probabilistic genotyping offers a way to utilize more of the data in a profile. Based on the studies performed in this validation, STRmix can be a powerful tool for analysis of mixtures representing one, two, and three individuals. Mixture profiles were analyzed using both STRmix and CPI, and STRmix provided a more discriminating statistic every time. The ability to use peak height data, incorporate dropout, and compare evidence to a reference profile makes STRmix a more comprehensive method of evaluation.

While STRmix proved useful in mixture analysis, there were limits to its function. Any extra peaks (drop-in, pull-up, etc.) over 75 rfu can cause an erroneous result. This is potentially problematic, especially for drop-in peaks, since they are nearly impossible to identify. Drop-in is rare, but analysts should carefully consider assigning a number of contributors if drop-in is suspected to avoid an incorrect result.

Results from four person mixtures, specifically from the minor contributors of these mixtures, were sometimes inconsistent with expected results, particularly when the minor contributor was present in low quantity. The major contributors of these mixtures did not always receive consistent LRs, as seen in the Precision study.

STRmix will likely be continually validated. Since the settings are partially based on peak heights and stutter ratios, a new validation must be performed for each different type of

49

amplification kit and CE instrument used for evidence samples. The software developers also periodically release updates to improve performance. With each update, some samples must be retested to ensure concordance with previous results. Through repeated validations for various laboratory methods as well as performance checks, the accurate operation of the software will be constantly substantiated.

REFERENCES

- (1) Kerr, Z. Two years later: A reflection on the implementation of STRmix in a high throughput DNA laboratory.
- (2) Coble M, Buckleton J, Butler HM, Egeland T, Fimmers R, Gill P, Gusmao L, Guttman B, Krawczak M, Morling N, Parson W, Pinto N, Schneider PM, Sherry ST, Willuweit S, Prinz M. DNA Commission of the International Society of Forensic Genetics: Recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications. *Forensic Sci Int Genet*. 25 (2016) 191-197.
- (3) Bieber FR, Buckleton JS, Budowle B, Butler JM, Coble MD. Evaluation of forensic DNA mixture evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion. *BMC Genetic*. 17 (2016) 1-15.
- (4) Institute of Environmental Science and Research Limited. STRmix[™] v2.3 User's Manual (2015).
- (5) Budowle B, Onorato AJ, Callaghan TF, Manna AD, Gross AM, Guerrieri RA, Luttman JC, McClure DL. Mixture Interpretation: Defining the Relevant Features for Guidelines for the Assessment of Mixed DNA Profiles in Forensic Casework. *J Forensic Sci.* 54 (2009) 810-821.
- (6) Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR, Morling N, Prinz M, Schneider PM, Weir BS. DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci Int.* 160 (2006) 90-101.

- (7) Bille TW, Weitz SM, Coble MD, Buckleton J, Bright J-A. Comparison of the performance of different models for the interpretation of low level mixed DNA profiles. *Electrophoresis*. 35 (2014) 3125–3133.
- (8) Curran J. A MCMC method for resolving two person mixtures. *Science and Justice*. 48 (2008) 168-177.
- (9) Bright J-A, Taylor D, McGovern C, Cooper S, Russell L, Abarno D, Buckleton J. Developmental validation of STRmix, expert software for the interpretation of forensic DNA profiles. *Forensic Sci Int Genet*. 23 (2016) 226-239.
- (10) Greenspoon SA, Schiermeier-Wood L, Jenkins BC. Establishing the Limits of TrueAllele Casework: A Validation Study. *J Forensic Sci.* 60 (2015) 1263-1276.
- (11) Scientific Working Group on DNA Analysis Methods (2015) Guidelines for the Validation of Probabilistic Genotyping Systems.
- (12) Institute of Environmental Science and Research Limited. STRmix[™] v2.3
 Implementation and Validation Guide (2014).
- (13) Balding DJ and Nichols RA. DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Sci Int*, 64 (1994) 125-140.